Structure and Dynamics of Phosphate Linkages and Sugars in an Abasic Hexaloop RNA Hairpin

Flore Joli,*† Edith Hantz,* and Brigitte Hartmann†

*Laboratoire BioMoCeTi, CNRS UMR 7033, UFR SMBH, Université Paris 13, Bobigny, France; and †Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-Chimique, Paris, France

ABSTRACT Hairpins containing hexaloops are well represented among the diverse conformations adopted by the RNA molecules. To investigate the intrinsic properties of a backbone submitted to a hexaloop fold, we present here a molecular dynamics study of an abasic hexaloop closed by an A-form 6 basepair stem. The analysis of the 23 ns trajectory made in explicit solvent shows that both the sugars and the torsion angles in the loop undergo numerous conformational transitions. The south sugars, although not in a majority, are the major actors of the loop stretching. The five torsion angles, ε, ζ, α, β, and γ, are unequally variable, and only ζ and α exhibit trimodal distributions. The analysis of the phosphate linkages in terms of ε-$\tilde{\zeta}$-$\tilde{\alpha}$-$\tilde{\beta}$-γ-combinations allows us to define five conformational families, each one composed of one major substate in equilibrium with several less populated ones. The transitions between the substates within a family follow specific pathways involving the angles ε, ζ, and α. Thus, this work reveals that the backbone conformational space is both reduced and ordered even in a hexaloop devoid of bases.

INTRODUCTION

RNA exhibits a large diversity of conformations and among them hairpin structures are frequently encountered. Hairpins are often involved in interaction with proteins and nucleic acids (1–5). It is tempting to postulate that loops are predisposed to various interactions because of the accessibility and the flexibility of the bases that are probably greater in this motif than in double helices. Along these lines, determination of the loop structures and dynamics in solution is especially important to better understand the biological functions of RNA. Each nucleotide is defined by seven parameters: the five torsion angles α, β, γ, ε, and ζ, the sugar pucker, and the torsion angle χ around the glycosidic bond. Although it is known that the elastic properties of the sugar-phosphate chains play a key role in folding shapes and stabilities of RNA hairpins (6), the multidimensionality of the nucleotide conformational space represents a major obstacle to acquiring reliable information about loops. The glycosidic angles and the sugar conformations can be estimated with sufficient precision by some characteristic distances and coupling constants measured by NMR (7). In contrast, α-, β-, γ-, ε-, and ζ-angle conformations are difficult to observe. In NMR, the interpretation of ^{31}P chemical shifts in terms of torsion angle conformations seems trustworthy only for B-DNA structures (8), in which only the correlated angles ζ and ε are variable (9). In a loop, there is no justification a priori for the maintenance of this ε/ζ-correlation; furthermore, we expect that α, β, and γ do not remain confined to one conformation. In this context, the ^{31}P chemical shifts are not yet interpretable. Traditionally, γ- and β-angle values can be estimated using the coupling constants $^3J_{H4'\text{-}H5'/H5''}$ and $^3J_{P\text{-}H5'/H5''}$, thanks to the Karplus relationship (10) and the intranucleotide distances d_i(H3′–H5′/H5′′) and d_i(H4′–H5′/H5′′). Unfortunately, the H5′/H5′′ attributions are often impossible, owing to strong spectral overlaps of these proton resonances. Finally, measurements of $^3J_{C4'\text{-}P}$ and $^3J_{C2'\text{-}P}$ on ^{13}C-labeled oligonucleotides provide information on the ε- and β-torsion angles (11). Nevertheless, as we will show, the β-angle is quasi-invariant, and, consequently, only the ε-determination is interesting. Thus, backbone conformations in RNA loops remain quite resistant to experimentation in solution.

To tackle the diversity of the backbone structures in RNA, using a steric criterion alone, the construction of a map of the conformations sterically accessible to ribonucleosides and truncated ribodinucleotides (12) suggested that the conformational space explored by the RNA backbone was limited. Recently, two fine analyses of large crystallographic datasets were published that confirmed this conclusion. Using quality filtering techniques, the analysis of L. J. Murray et al. (13) allowed the identification of 42 conformational combinations for the ribose-to-ribose backbones (δ_i-ε-ζ-α-β-γ-δ_{i+1}, so-called ''suites'' by the authors). B. Schneider et al. (14) showed, on 3000 nucleotides of ribosomal RNA, that α and ζ were the most variable angles in the backbone and detected 32 non-A and A type conformations for the ribose-to-ribose units. However, these articles did not distinguish the loops from other structures. The loop backbones in hairpins are subjected to closing constraints and we could guess that the associated conformational space is reduced. It seems to be the case for tetraloops that show numerous common structural features (15–17).

Submitted July 15, 2005, and accepted for publication October 12, 2005.

Address reprint requests to Flore Joli, Laboratoire BioMoCeTi, CNRS UMR 7033, UFR SMBH, Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny cedex, France. E-mail: fjoli@smbh.univ-paris13.fr.

0006-3495/06/03/1480/09 $2.00

doi: 10.1529/biophysj.105.070862

In this work, we investigate the intrinsic properties of an abasic hexaloop by analyzing a 23 ns trajectory in explicit solvent. Removing the bases in the loop reduced it to the sugars and phosphate groups, allowing us to focus on the structure and dynamics of the backbone. This abasic loop is closed by a stem (18) containing a G·U mismatch that is part of the mRNA of the PGY1/MDR1 gene encoding the transmembrane P-glycoprotein (P-gp). The NMR data collected on the native hairpin highlighted numerous C2′-*endo* sugars and marked up- and down-field phosphorus chemical shifts in the GGGAUG loop (19). The analysis of the simulation made on the abasic loop shows that both the sugars and the backbone angles undergo frequent conformational transitions. However, the classification of the phosphate linkages in terms of $\varepsilon/\zeta/\alpha/\beta/\gamma$-conformer suites allows us to define only five conformational families, each one containing from three to six substates in equilibrium. Thus, the multidimensional conformational space explored by the backbone can be sorted into preferred, allowed, and disallowed regions; furthermore, the relationships between the substates will be discussed in terms of transition pathways.

METHODS

Hairpin building

The model was built and minimized with the internal coordinate program JUMNA (20,21). We started from a first strand composed of the 5′-*r*(GAGGUCOOOOOO)-3′ sequence (''O'' represents an abasic site in which the base was replaced by a proton (see Fig. 2)) and a second strand containing the 5′-*r*(GAUCUC)-3′ sequence. The six bases of the two strands formed an A-RNA double helix, whereas the abasic part was an extended single strand. Then, constrained distances were applied by steps of 0.2 Å between the O3′ atom of the last O-3′ site of the first strand and the C5′ atom of the 5′-G of the second strand. At each step, the structure was minimized until the abasic part was folded in a loop. After a last minimization in JUMNA, the resulting structure was used in the molecular dynamics (MD) program.

Molecular dynamics

The simulations were performed using the AMBER 7.0 program (22) and the Parm99 force field (23). An MD simulation of 23 ns was carried out on the hairpin neutralized with 17 Na^+ counterions (one Na^+ for each phosphate group) and explicitly solvated by a 12 Å water shell in all directions (5641 TIP3P water molecules) in a truncated octahedral box of face-to-face dimensions of ~75 Å. After 2250 cycles of energy minimization, the minimized system was heated to 300 K, rescaling the velocities as necessary, and coupling to a heat bath using the Berendsen algorithm (24). During each of these phases, harmonic restraints were imposed on the atomic positions of the oligomers and then slowly relaxed over several periods of 50 ps until a free system was achieved. The simulations were then performed at constant temperature and pressure using the Berendsen algorithm. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm (25), which enabled an integration time step of 2 fs. Long-range electrostatic interactions were treated using a periodic system, i.e., the particle mesh Ewald approach (26,27). The Lennard-Jones interactions were cut off at a distance of 9 Å. The direct sum tolerance criterion was 10^{-5}, and the reciprocal space charge grid spacing was ~1 Å. During the 23 ns of the production phase, translations and rotations of the stem-loop oligomer were removed every 100 steps.

Structural analysis

The oligomer residues were numbered as indicated in Fig. 1, the abasic sites being noted ''O''. Our analysis focused on the behavior of the five phosphodiester linkages (from O7pO8 to O11pO12) belonging to the loop. In this article, we describe the backbone arrangement of a given dinucleotide by the following series of torsion angles $_{5'}\varepsilon$-ζ-α-β-$\gamma_{3'}$, with ε:C4′-C3′-O3′-P; ζ:C3′-O3′-P-O5′; α:O3′-P-O5′-C5′; β:P-O5′-C5′-C4′; and γ:O5′-C5′-C4′-C3′. This suite is illustrated in Fig. 2. The torsion angles are described by the classical threefold staggered pattern of the dihedrals: g^+ (60° ± 60°), *trans* (180° ± 60°), and g^- (300° ± 60°). The pseudorotation space is divided into four equally sized quadrants centered around $P = 0°$, $P = 90°$, $P = 180°$, and $P = 270°$ that are referred to as the north (also called C3′-*endo*), east, south (also called C2′-*endo*), and west quadrants, respectively.

In addition to the examination of the root mean-square deviation (RMSD) time series, the convergence of the simulations is tested by comparing the behavior of the identical abasic residues occurring in the loop. For example, the average values of the puckers of the loop sugars become similar only after 3 ns. Consequently, only the last 20 ns of the simulation was used for the analysis.

RESULTS AND DISCUSSION

General characteristics of the molecular dynamics simulation

The RMSD between the starting structure and the hairpin snapshots is 3.5 ± 0.6 Å in average. The RMSD for the average structure is 2.2 ± 0.4 Å; this value ensures that a large conformational space is explored without going toward extreme exotic structures (see also Fig. 3). During the MD, the six basepair stems stay close to a regular A conformation, with an average RMSD of 1.35 ± 0.1 Å (Fig. 4), the RMSD

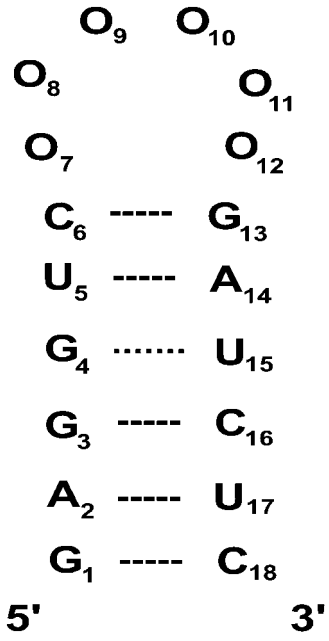

FIGURE 1 Numbering of the RNA hairpin. O, abasic residue.

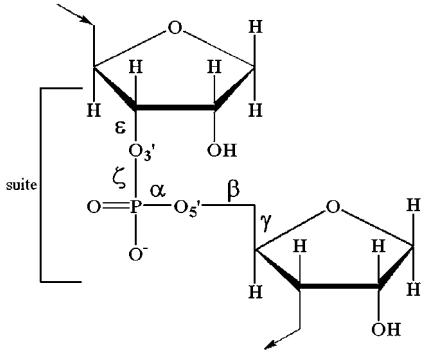

FIGURE 2 RNA backbone, with the five torsion angles ε, ζ, α, β, and γ (called ''suites''). Note that the sugars contain two hydrogens on the C1′ atom, as in the abasic loop.

with respect to the average structure being 0.75 ± 0.2 Å. The two strands are connected by stable hydrogen bonds, comprising the G·U mismatch that, as described elsewhere (19,28–30), adopts a ''wobble'' conformation (hydrogen bonds between N3(U) and O6(G) and between O2(U) and N1(G)). In contrast, the loop part clearly exhibits higher RMSD. We obtained RMSDs of 4.0 ± 0.6 Å (Fig. 4) taking the starting point as a reference, and 2.8 ± 0.4 Å with respect to the average structure. This region clearly explores a larger conformational space than the stem.

Sugar conformations

Eleven of the twelve sugars located in the stem populate the C3′-*endo* conformation with an average *P* value of 10.15° ± 20°. Only the sugar belonging to the U5 base in the 3′ side of the mismatch undergoes C2′-*endo*/C3′-*endo* transitions.

The sugars in the loop are all submitted to high-frequency transitions between three conformational regions: the majority adopt (79%) the north conformation, but the south and west regions are also populated. Nevertheless, the west sector does not seem as stable as the north and the south sectors: the lifetimes of west sugars are <10 ps, in contrast with several hundred picoseconds for the north and south lifetimes (Fig. 5). On average, the loop contains three sugars in a non-north conformation. The presence of south sugars is not surprising, given the small south/north energy barriers (≤0.5 kcal/mol) obtained on ribonucleosides by high-level quantum mechanical calculations in vacuo (31) or with the AMBER and CHARMM27 force fields in explicit solvent (32). Furthermore, our analysis of NMR structures extracted

FIGURE 3 Eight MD average structures. The superimpositions were made on the stem parts of the hairpin. The chosen structures are the average ones, calculated on the eight clusters identified on a cross-RMSD map using a cutoff of 2.8 Å.

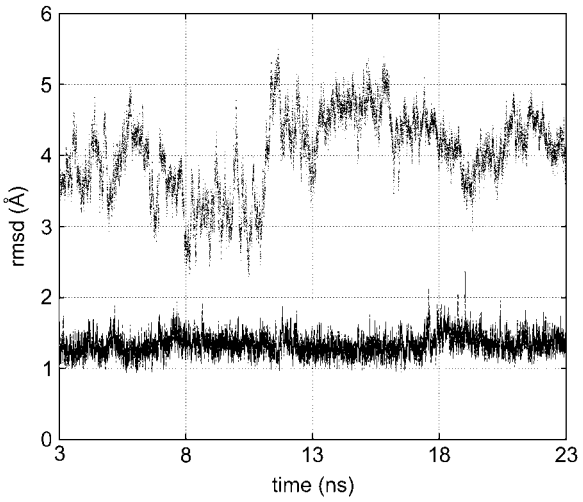

FIGURE 4 Evolution of the RMSD between the initial structure and the MD snapshots during the trajectory: solid line, the stem; dashed line, the abasic loop.

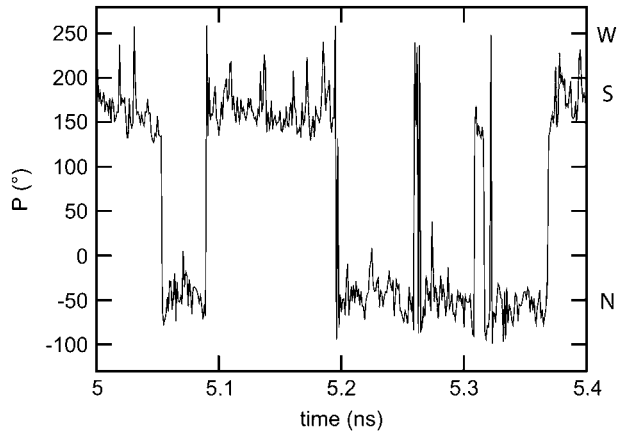

FIGURE 5 Fluctuations of the sugar 10 pucker (*P*) in the loop during a part of the simulation to illustrate the explored conformations (*W*, west; *S*, south; *N*, north) and their lifetimes.

from the hexaloop database (33) shows that south sugars are commonly encountered in loops. For some of these loops, homonuclear three-bond $^3J_{H1',H2'}$ coupling constant measurements (34,35) ensure that the presence of south sugars is not due to a systematic force-field bias in refinements. In contrast, the west conformation is usually considered improbable (36). Here, we cannot exclude an undesirable effect of the force field: N. Foloppe and L. Nilsson (32) reported a significant percentage (~4%) of west sugars in nucleosides minimized with AMBER that are not retrieved with CHARMM27. However, strange and very rare conformations of sugars could be due to the absence of bases in the loop, conferring a large degree of freedom. Indeed, both south and west sugars contribute to expand the loop backbone by 0.5 Å on average, as previously noticed by E. J. Sorin et al. (37) for the south sugars: the values found for $D_{O5'-O3'}$, the distance between the O5′ and O3′ atoms enclosing the considered sugar, are 4, 4.2, 4.5, and 4.55 Å for north, east, south and west sugars, respectively.

ε-ζ-α-β-γ-conformational combinations

In contrast with the stem, in which the ε-, ζ-, α-, β-, and γ-angles are locked in canonical conformations (ε: *trans*; ζ: g^-; α: g^-; β: *trans*; and γ: g^+), these backbone angles in the loop are able to adopt a wide range of conformations. The value distributions for each angle in the abasic part are shown in Fig. 6. The angles ζ and α are the most variable: they populate the entire dihedral space, although the *trans* region appears disfavored in comparison with the g^- and the g^+ regions. The ε- and γ-distributions are bimodal, the g^+ and g^- regions being disallowed by ε and γ, respectively. The torsion angle β-distribution exhibits a unique peak centered at 180°. Despite the absence of bases, these distributions are totally in

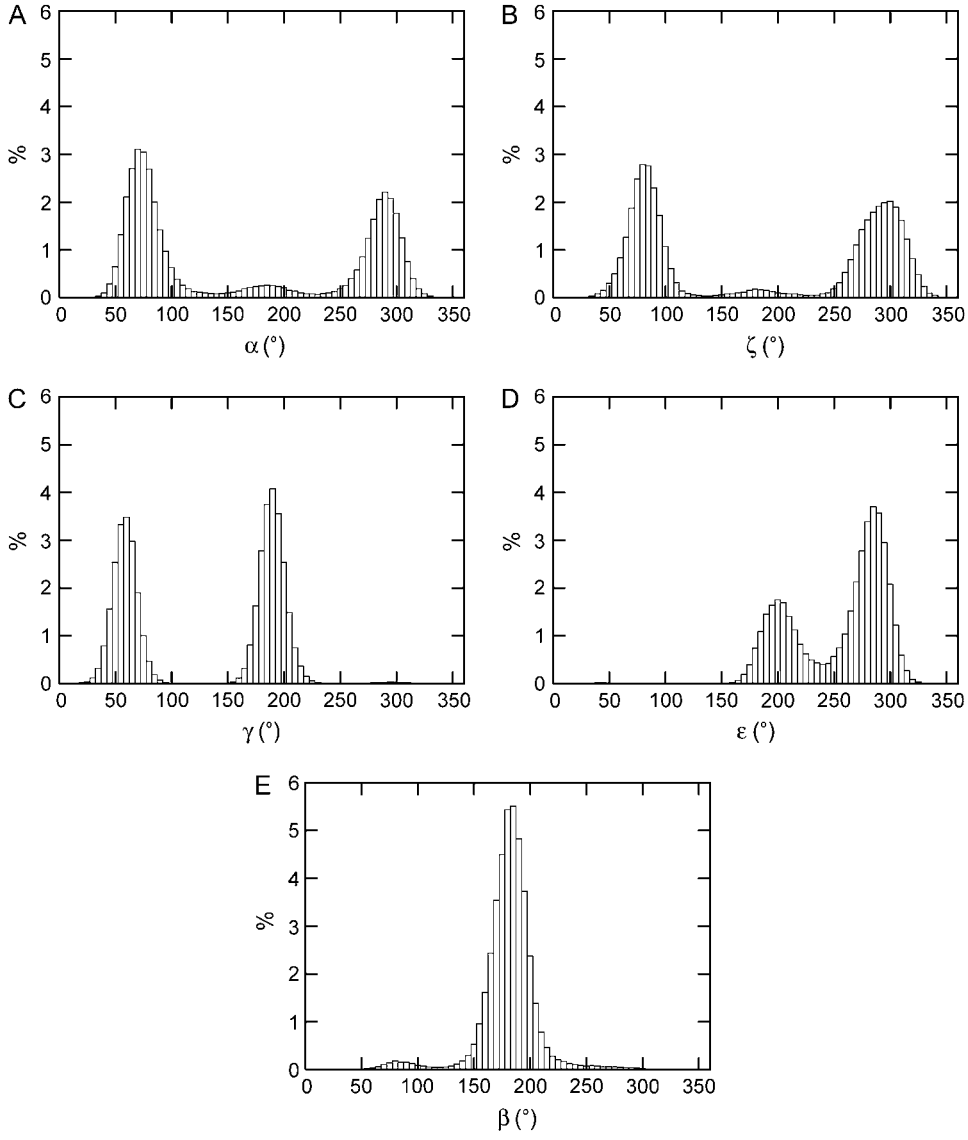

FIGURE 6 Distribution (%) of the torsion angles α- (*A*), ζ- (*B*), γ- (*C*), ε- (*D*), and β- (*E*) values in the abasic loop.

agreement with two previous backbone analyses made on large RNA crystallographic structures with bases (13,14).

Apart from the β-angle, the percentages of noncanonical conformations in the abasic steps lie between 52% (for ζ) and 65% (for ε) (Table 1). The loop cannot exist with five canonical ε-ζ-α-β-γ-suites. The fold necessitates on average the simultaneous occurrence of three noncanonical ε-, ζ-, and α-angles and between two and four noncanonical γ-angles. Overall, half (12/25) of the angles composing the five abasic phosphodiester linkages adopt noncanonical conformations. Parsing the abasic steps in terms of ε-ζ-α-β-γ-suites from sugar_i to sugar_{i+1}, 21 conformational substates (Table 1) among the 36 possible combinations ($3 \times 3 \times 2 \times 2$, following the distributions observed above) are populated at least 1% of the time. These substates cover 92% of the structures, the remaining ones being distributed into classes containing $<1\%$ of the 100,000 phosphodiester linkages we examined. As the west sugars are subject to caution, this clustering was also made on structures devoid of this type of sugar: this leads to the same result, i.e., the same categories in the same order.

This list of 21 combinations can be revisited as a function of the dynamical relations that we observe between the substates. Thus, assembling the substates that can easily coexist in equilibrium, the 21 combinations are categorized into only five families (Table 1). In this classification, we first take into account that half of the ζ-α-β-γ-combinations support ε either in *trans* or in g^- conformations. Although the ε:*trans* and g^- conformations only slightly overlap (Fig. 6 *D*), this angle undergoes frequent transitions between these two regions, as shown in Fig. 7 *A* for substates 1 and 4. This observation explains, in terms of dynamical behavior, a previous analysis of crystallographic structures made on RNA with bases present (13) that led the authors to combine ε:*trans* and ε:g^- in a so-called ''eclipsed'' conformation. Second, we consider more complex equilibriums implying α- or ζ-transitions in addition to the ε one. One example is given in Fig. 7 *B* for the four substates 7, 8, 13, and 16 that are in equilibrium via both ε- and ζ-transitions. Finally, it should be noted that the γ-angle is not useful to define families, as it never displays any fast transition between its two conformers g^+ and *trans*, the lifetimes of each conformer attaining several nanoseconds. This strong stability is reflected by the clear separation of the distribution peaks (Fig. 6 *C*), moreover affected by the smallest standard deviations (11° vs. 16° on average for the ε-, ζ-, α-, and β-conformers).

Each family is characterized by lifetimes of several nanoseconds. In contrast, within a given family, the minor substates ($\sim$2%) are clearly more fleeting, as illustrated in Fig. 7 *A* for substates 10 and 11, belonging to family I. Nevertheless, in some relatively rare cases, these minor substates can appear alone during a few hundred picoseconds: Fig. 7 *C* shows substate 21 enclosing the stable substate 7, which does not belong to the same family.

Several different pathways are observed between the families themselves. Similarly, the substates in families IV and V exchange without any particular trail. The situation is completely different within families I, II, and III, in which

TABLE 1 Characterization of the families and the substates found for the ε-ζ-α-β-γ-abasic loop suites

Family	Substate	ε	ζ	α	β	γ	%	5′-P	P-3′	Schneider's classes
I	**1**	$\boldsymbol{g^-}$	$\boldsymbol{g^+}$	$\boldsymbol{g^+}$	$\boldsymbol{t}$	$\boldsymbol{t}$	**21.5**			**Non-A**
	4	t	g^+	g^+	t	t	7	*N*		Non-A
	10	g^-	g^+	t	t	t	2			
	11	t	t	g^+	t	t	2			
II	**2**	$\boldsymbol{g^-}$	$\boldsymbol{g^-}$	$\boldsymbol{g^-}$	$\boldsymbol{t}$	$\boldsymbol{g^+}$	**17**			**A**
	5	t	g^-	g^-	t	g^+	7			Canonical A
	20	t	g^+	g^-	t	g^+	1–2	*N*	*N*	
	21	t	g^-	t	t	g^+	1–2	*N*		A/non-A
III	9	g^-	g^-	g^-	t	t	2.5		*N*	
	3	$\boldsymbol{t}$	$\boldsymbol{g^-}$	$\boldsymbol{g^-}$	$\boldsymbol{t}$	$\boldsymbol{t}$	**9**	***N***		**A**
	12	t	g^-	t	t	t	1–2	*N*		A
	14	g^-	g^-	g^+	t	t	1–2			
	17	g^-	t	g^-	t	t	1–2			
	18	t	t	g^-	t	t	1–2	*N*		
IV	**6**	$\boldsymbol{g^-}$	$\boldsymbol{g^+}$	$\boldsymbol{g^+}$	$\boldsymbol{t}$	$\boldsymbol{g^+}$	**5**			**Non-A**
	19	t	g^+	g^-	t	g^+	1–2			Non-A
	15	g^-	g^+	t	t	g^+	1–2		*N*	
V	**7**	$\boldsymbol{g}$-	$\boldsymbol{g^-}$	$\boldsymbol{g^+}$	$\boldsymbol{t}$	$\boldsymbol{g^+}$	**4.5**			**Non-A**
	16	t	g^-	g^-	t	g^+	1–2	*N*	*N*	
	8	g^-	t	g^+	t	g^+	3.5		*N*	
	13	t	t	g^+	t	g^+	1–2		*N*	Non-A

The numbering of the substates corresponds to the decreasing order of the occurrences (%). The major substate within a given family is in bold. In columns 9 and 10, *N* indicates that the north sugar percentage in 5′ (5′-P) or in 3′ (P-3′) sugars is $\geq$90% within the corresponding substate. The last column refers to the families described in Schneider et al. (14).

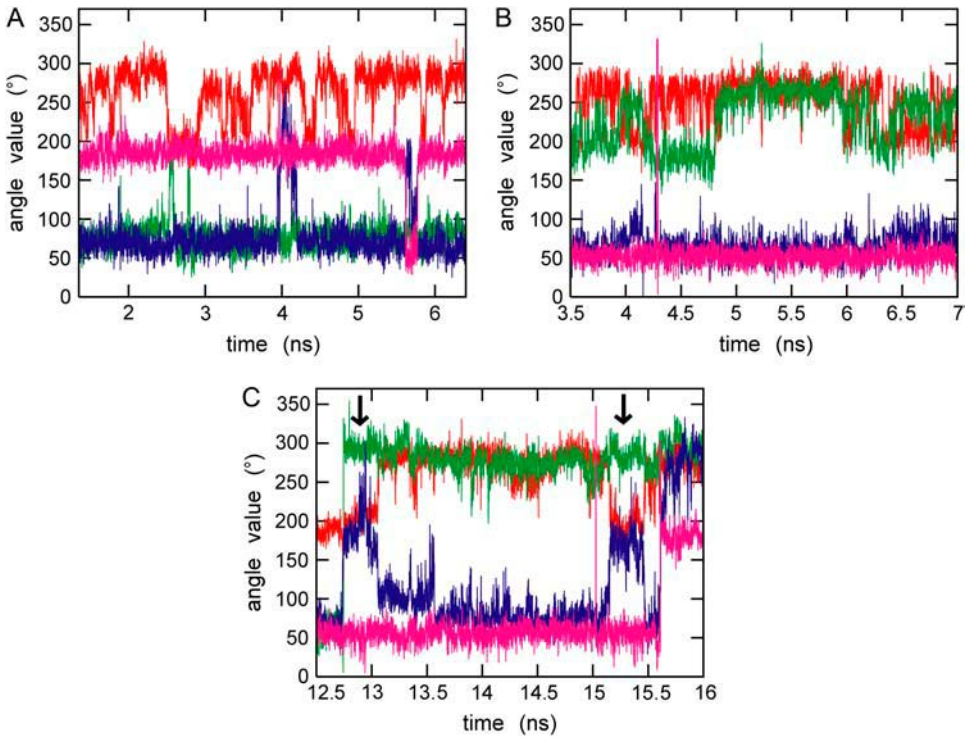

FIGURE 7 ε- (*red*), ζ- (*green*), α- (*blue*), and γ- (*pink*) fluctuations during parts of the simulation for (*A*) the step 9p10 (equilibrium between the substates 1, 4, 10, and 11), (*B*) the step 11p12 (equilibrium between the substates 7, 8, 13, and 16), and (*C*) the step 9p10 (substate 21 is indicated by the *arrows*). The definitions of the substates are given in Table 1.

the substates are in equilibrium through the following pathways (the substates noted (N/M) indicate a fast ε-equilibrium):

$$10 \leftrightarrow (4/1) \leftrightarrow 11$$
$$20 \leftrightarrow (5/2) \leftrightarrow 21$$
$$(18 \leftrightarrow 17) \leftrightarrow (3/9) \leftrightarrow 12 \leftrightarrow 14.$$

These transition pathways are visible in ζ/α-2D representations plotted for γ:g^+ (Fig. 8, *A* and *B*) and γ:*trans* (Fig. 8, *C* and *D*). We observe furthermore that the ζ/α:*trans*/*trans* combination is almost totally forbidden and that the quadrant ζ/α:g^+/g^- is only populated by substate 20. The quadrant ε/ζ/α/β/γ:*trans*/*trans*/g^-/*trans*/g^+, poorly populated (0.5% of the snapshots), is nevertheless the passage between substate 20 and substates 5 or 2.

These substates could be typical of abasic hexaloop and thus not in position to be extrapolated to any other RNA structure. So, the comparison with the experimental data obtained on ''real'' RNA structures (with bases) is pertinent. Unfortunately, the available solution experimental data extracted on nonabasic loops are not conclusive, as discussed in the introduction. The situation is totally different for RNA crystallographic data, on which several analyses were made, especially on ribosomal RNA. The analysis of the HM LSU 23S rRNA (17) concentrated only on torsions α, γ, and ζ, and thus the backbone conformations were not described sufficiently to be compared with our substates. In contrast, L. J. Murray et al. (13) and B. Schneider et al. (14) provided series of families extracted from non-A- and A-type conformations, all of them containing bases, and based on the backbone conformations, comprising the sugar conformations ($P_{5'}$-ε-ζ-α-β-γ-$P_{3'}$ combinations). In our case, non-north sugars do not correspond to any particular torsion suite: in other words, most of the ε-ζ-α-β-γ-combinations clearly support either non-north or north sugars, whereas a minority do not. The only link between the sugar and the angle conformations that can be found everywhere is the noticeable preference of the 5′ south sugar for ε-g^- (Table 1), also reported in the two articles of interest. When the lists are strictly based on the ε-, ζ-, α-, β-, and γ-conformers, the 42 combinations found by Murray et al. and the 32 described by Schneider et al. are reduced to 14 and 13, respectively. Almost all of our classes are retrieved in Murray's families, but this evaluation is slightly biased by the fact that the authors do not distinguish ε:*trans* and ε:g^-. As Schneider et al. do not make this simplification, their substates allow a finer comparison. Eleven families are strictly identical in the two analyses. They belong to our most representative substates, as they include our first seven. Schneider's equivalents of seven substates are listed in the non-A type conformations without stacking, whereas four substates appear in the A-RNA type conformations (Table 1). As some of these substates are weakly populated in our simulations, it emerges from this comparison that the presence of bases can stabilize them. The fact that two substates present in Schneider's list are absent in ours could indicate a default in the MD sampling. This is not the case, however, as they are found (each

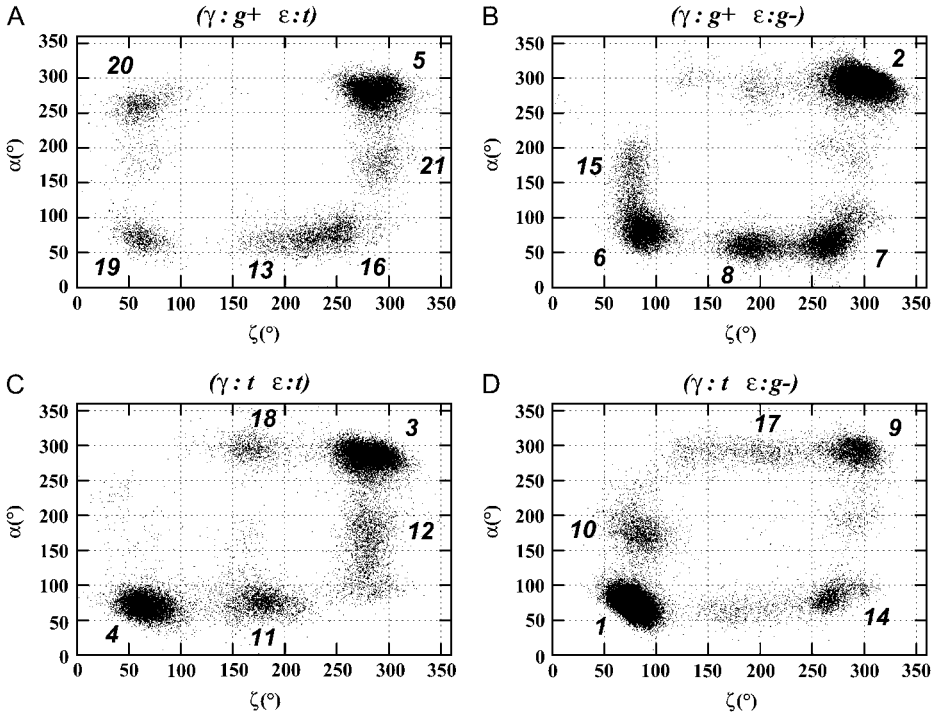

FIGURE 8 Scattergrams of ζ versus α for γ:g^+/ε:*trans* (*A*); γ:g^+/ε:g^{--} (*B*); γ:*trans*/ε:*trans* (*C*), and γ:*trans*/ε:g^- (*D*) in the abasic loop. The numbers beside the dots refer to the substates defined in Table 1.

~0.25%) among the remaining MD structures that do not appear in Table 1. Finally, the nine substates without equivalents in Schneider's list correspond to small populations and they could be considered transient substates rather than stable states. Nevertheless, in these cases, it is difficult to know whether they are specific to an abasic hexaloop or whether they are simply not trapped in the crystallographic structures.

Characteristic distances of the ε-ζ-α-β-γ-loop substates

According to the substates, the average $C3'_i$–$C4'_{i+1}$ distances, i and $i + 1$ labeling two successive sugars ($D_{C3'C4'}$, Table 2), excluding the $C4'_i$–$C3'_i$ and $C4'_{i+1}$–$C3'_{i+1}$ sugar linkages, vary between 4.7 and 5.7 Å (5.0 Å for the canonical conformation). Taking into account the standard deviations (0.4 Å whatever the substate), some combinations of non-canonical angles seem able to extend the loop. Nevertheless, these distances appear homogeneous within a given family, and those of the first families (from I to III, i.e., the most populated) oscillate around 5 Å, the characteristic distance of the canonical conformation. Families IV and V contain two substates that clearly differ from the canonical one, but, overall, these four classes only represent 8% of the observed backbones. As seen above, each non-north sugar stretches the loop by 0.5 Å, and, as their ratio reaches 21%, it should be concluded that the major actors of the loop extension are the sugars rather than the ε-ζ-α-β-γ-combinations.

The $C1'_i$–$C1'_{i+1}$ distances ($D_{C1'}$) are related to the orientation of two successive sugars, approximated by the angle between two successive C4′-C1′ vectors ($V_{C4'\text{-}C1'}$). The $V_{C4'\text{-}C1'}$ distribution (not shown) reveals four well defined peaks that correspond to parallel ($V_{C4'\text{-}C1'}$ ~0°), antiparallel ($V_{C4'\text{-}C1'}$ ~180°), and perpendicular ($V_{C4'\text{-}C1'}$ ~±90°) sugars. As a reference, $V_{C4'\text{-}C1'}$ is 24° in a duplex in canonical A-form, a value that is retrieved in our stem, with a standard deviation of ±8°. When the sugars are parallel, $D_{C1'}$ is minimal (5.2 ± 0.5 Å in our loop, 5 Å in a canonical A-RNA), whereas it is maximal (8.3 ± 1.6 Å) for antiparallel sugars. In the majority (62%) of the sugar-to-sugar conformations, the sugars are perpendicular ($V_{C4'\text{-}C1'}$ ~±90°) and are associated with intermediate $D_{C1'}$ values (6 ± 0.8 Å). These parameters do not really characterize any substate since, generally, two successive sugars within a substate can adopt any relative position. Indeed, these positions are influenced by both the combinations of the flanking steps and the substates with which they undergo equilibrium. Nevertheless, some trends can be identified: substates 3, 10, 11, and 12 are almost devoid of antiparallel sugars, whereas parallel sugars represent <10% of the numerous substates (2,4,15,18–21).

Hydrogen bonds within the loop

The distance between the HO2′ and O3′ atoms is from 1.6 to 3.9 Å. When these atoms are close to each other (<2.5 Å), numerous hydrogen bonds occur between them during the simulations, despite the explicit presence of water molecules. Taking 80° as the upper limit for the valence angle HO2′-O2′-O3′, we find that these hydrogen bonds are present 55% and 41% of the time in the loop and in the stem, respectively, with lifetimes rarely exceeding 50 ps. The ε:g^- conformer makes the HO2′–O3′ and O2′–O3′ distances slightly

TABLE 2

Family	Substate	ε	ζ	α	β	γ	$D_{C3'-C4'}$ (Å)	$D_{C1'}$ (Å)
I	1	g^-	g^+	g^+	t	t	5.1	6.7
	4	t	g^+	g^+	t	t	4.9	6.7
	10	g^-	g^+	t	t	t	5.2	6.4
	11	t	t	g^+	t	t	5.0	5.2
II	2	g^-	g^-	g^-	t	g^+	5.0	7.2
	5	***t***	***g*$^-$**	***g*$^-$**	***t***	***g*$^+$**	**5.0**	**6.9**
	20	t	g^+	g^-	t	g^+	4.8	7.8
	21	t	g^-	t	t	g^+	5.1	6.4
III	9	g^-	g^-	g^-	t	t	4.7	6.5
	3	t	g^-	g^-	t	t	5.0	7
	12	t	g^-	t	t	t	5.0	6
	14	g^-	g^-	g^+	t	t	5.1	6.6
	17	g^-	t	g^-	t	t	5.3	7.6
	18	t	t	g^-	t	t	5.0	7.2
IV	6	g^-	g^+	g^+	t	g^+	5.2	6.4
	19	t	g^+	g^+	t	g^+	5.7	6.8
	15	g^-	g^+	t	t	g^+	5.4	6.6
V	7	g^-	g^-	g^+	t	g^+	4.9	7.2
	16	t	g^-	g^+	t	g^+	4.9	6.8
	8	g^-	t	g^+	t	g^+	5.4	7.7
	13	t	t	g^+	t	g^+	5.4	7.1

For clarity, Table 2 reproduces a part of Table 1. The last two columns list the distances between the C3′ atom of a sugar i and the $C4'_{i+1}$ atom of the successive sugar $i + 1$ ($D_{C3'-C4'}$) and between two C1′ successive sugar atoms ($D_{C1'}$). The standard deviations are given in the text.

shorter, by 0.3 Å, in regard to the ε:*trans* conformer, and thus favors hydrogen bonding. This fact is reflected in the difference between the hydrogen bond occurrences within the loop and the stem, in which ε is locked in *trans* conformation. In contrast, an examination of the few loop backbones containing an angle ε:g^+ (0.34% of the snapshots) shows that the distance O2′–O3′ becomes >3 Å, so that no hydrogen bond is stable any more. It could be the reason why, although this conformation is possible in terms of fold, it is almost never encountered in this type of structure.

CONCLUSION

This modeling study aims to investigate the intrinsic dynamical properties of the RNA backbone submitted to a hexaloop fold. The use of a hairpin containing an abasic loop allows us to release the effects of base sequences which, through stacking patterns and, sometimes, interbase hydrogen bonds, restrict the conformational space accessible to the backbone. A major limitation of the modeling methods is the quality of the sampling. Despite the fact that we made a rather long simulation of 23 ns, it is difficult to ensure the complete convergence of our system. Nevertheless, we confirmed that, apart from angle β, classically frozen in the *trans* conformation, all the torsion angles explore the three possible conformations, even if some of them do not appear in our classification on account of their instability. However, the percentages of the substates presented in Table 1 should be considered with some caution, especially when their lifetimes exceed several nanoseconds and do not warrant a large number of observations. Overall, the excellent consistency between our results and the previous studies devoted to analysis of crystallographic structures guarantees that the highlighted families make sense.

Our results show that the loop backbone expressed in terms of ε-ζ-α-β-γ-suites clearly has a preference for some regions of the multidimensional conformational space. This preference permits us to define a limited number of substates significantly but unequally populated. Furthermore, the MD method allows us to detect favored relationships between substates that lead to identify five families, each one composed of a major substate in equilibrium with several minor ones. Regarding the length of the backbone, we find that south sugars can stretch it more efficiently than the torsion angles. The ε-ζ-α-β-γ-combinations affect the relative position of two successive sugar planes. Finally, a comparison with previous classifications made on the basis of large crystallographic datasets, in particular those of Murray et al. (13) and Schneider et al. (14), shows a large overlap between our abasic loop substates and those detected on RNA with bases. Thus, our substates have a sense for any ''real'' RNA hexaloop. However, it should be underlined that the relative occurrences of the different substates cannot be easily extrapolated, as we expect that the presence of bases introduces many other energetic components influencing both the equilibrium and the structures of the backbone. Nevertheless, we believe and hope that knowledge of the intrinsic properties of the backbone can help in refining experimental x-ray and NMR structures.

We thank Christophe Oguey, Brahim Heddi and Nadia Bouchemal for helpful discussions, and Richard Lavery for a generous allocation of computer time.

REFERENCES

1. Legault, P., J. Li, J. Mogridge, L. E. Kay, and J. Greenblatt. 1998. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell.* 93:289–299.
2. Rijnbrand, R., V. Thiviyanathan, K. Kaluarachchi, S. M. Lemon, and D. G. Gorenstein. 2004. Mutational and structural analysis of stem-loop IIIC of the hepatitis C virus and GB virus B internal ribosome entry sites. *J. Mol. Biol.* 343:805–817.
3. Brunel, C., R. Marquet, P. Romby, and C. Ehresmann. 2002. RNA loop-loop interactions as dynamic functional motifs. *Biochimie.* 84: 925–944.
4. Mir, K. U., and E. M. Southern. 1999. Determining the influence of structure on hybridization using oligonucleotide arrays. *Nat. Biotechnol.* 17:788–792.
5. Shi, H., B. E. Hoffman, and J. T. Lis. 1997. A specific RNA hairpin loop structure binds the RNA recognition motifs of the *Drosophila* SR protein B52. *Mol. Cell. Biol.* 17:2649–2657.
6. Santini, G. P., C. Pakleza, and J. A. Cognet. 2003. DNA tri- and tetra-loops and RNA tetra-loops hairpins fold as elastic biopolymer chains in agreement with PDB coordinates. *Nucleic Acids Res.* 31:1086–1096.
7. Wijmenga, S. S., and B. N. M. van Buuren. 1998. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* 32:287–387.
8. Gorenstein, D. G. 1992. 31P NMR of DNA. *Methods Enzymol.* 211: 254–286.
9. Djuranovic, D., and B. Hartmann. 2003. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* 20:771–788.
10. Lankhorst, P. P., C. A. Haasnoot, C. Erkelens, and C. Altona. 1984. Carbon-13 NMR in conformational analysis of nucleic acid fragments. 2. A reparametrization of the Karplus equation for vicinal NMR coupling constants in CCOP and HCOP fragments. *J. Biomol. Struct. Dyn.* 1:1387–1405.
11. O'Neil-Cabello, E., Z. Wu, D. L. Bryce, E. P. Nikonowicz, and A. Bax. 2004. Enhanced spectral resolution in RNA HCP spectra for measurement of (3)J(C2′P) and (3)J(C4′P) couplings and (31)P chemical shift changes upon weak alignment. *J. Biomol. NMR.* 30:61–70.
12. Murthy, V. L., R. Srinivasan, D. E. Draper, and G. D. Rose. 1999. A complete conformational map for RNA. *J. Mol. Biol.* 291:313–327.
13. Murray, L. J., W. B. Arendall 3rd, D. C. Richardson, and J. S. Richardson. 2003. RNA backbone is rotameric. *Proc. Natl. Acad. Sci. USA.* 100:13904–13909.
14. Schneider, B., Z. Moravek, and H. M. Berman. 2004. RNA conformational classes. *Nucleic Acids Res.* 32:1666–1677.
15. Heus, H. A., and A. Pardi. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science.* 253:191–194.
16. Maier, A., H. Sklenar, H. F. Kratky, A. Renner, and P. Schuster. 1999. Force field based conformational analysis of RNA structural motifs: GNRA tetraloops and their pyrimidine relatives. *Eur. Biophys. J.* 28: 564–573.
17. Hershkovitz, E., E. Tannenbaum, S. B. Howerton, A. Sheth, A. Tannenbaum, and L. D. Williams. 2003. Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res.* 31:6249–6257.
18. Kostenko, E. V., R. S. Beabealashvilly, V. V. Vlassov, and M. A. Zenkova. 2000. Secondary structure of the 5′-region of PGY1/MDR1 mRNA. *FEBS Lett.* 475:181–186.
19. Joli, F., N. Bouchemal, B. Hartmann, and E. Hantz. 2006. NMR and molecular modelling studies of an RNA hairpin containing a G-rich hexaloop. *C. R. Chimie 9*. ISSN:1631–0748.
20. Lavery, R., K. Zakrzewska, and H. Sklenar. 1995. JUMNA (junction minimisation of nucleic acids). *Comput. Phys. Commun.* 91:135–158.
21. Varnai, P., and K. Zakrzewska. 2004. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* 32:4269–4280.
22. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, I. J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. Radmer, Y. Duan, J. Pitera, I. Massova, G. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman. 2002. AMBER 7. University of California, San Francisco.
23. Cheatham, T. E. III, P. Cieplak and P. A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* 16:845–862.
24. Berendsen, H. J. C., J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
25. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23: 327–341.
26. Darden, T., D. York, and L. J. Pedersen. 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
27. Cheatham III, T. E., J. L. Miller, T. Fox, T. Darden, and P. A. Kollman. 1995. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* 117:4193–4194.
28. McDowell, J. A., L. He, X. Chen, and D. H. Turner. 1997. Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of (rGGAGUUCC)2 and (rGGAUGUCC)2. *Biochemistry.* 36:8030–8038.
29. Varani, G., and W. H. McClain. 2000. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1:18–23.
30. Masquida, B., and E. Westhof. 2000. On the wobble GoU and related pairs. *RNA.* 6:9–15.
31. Foloppe, N., and A. D. MacKerell, Jr. 1999. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys. J.* 76: 3206–3218.
32. Foloppe, N., and L. Nilsson. 2005. Toward a full characterization of nucleic acid components in aqueous solution: similations of nucleosides. *J. Phys. Chem. B.* 109:9119–9131.
33. Tamura, M., D. K. Hendrix, P. S. Klosterman, N. R. Schimmelman, S. E. Brenner, and S. R. Holbrook. 2004. SCOR: structural classification of RNA, version 2.0. *Nucleic Acids Res.* 32:D182–D184.
34. Addess, K. J., J. P. Basilion, R. D. Klausner, T. A. Rouault, and A. Pardi. 1997. Structure and dynamics of the iron responsive element RNA: implications for binding of the RNA by iron regulatory binding proteins. *J. Mol. Biol.* 274:72–83.
35. Flodell, S., J. Schleucher, J. Cromsigt, H. Ippel, K. Kidd-Ljunggren, and S. Wijmenga. 2002. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* 30:4803–4811.
36. Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York.
37. Sorin, E. J., M. A. Engelhardt, D. Herschlag, and V. S. Pande. 2002. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* 317:493–506.